

## Л И Т Е Р А Т У Р А

1. Бакалов В. П., Дмитрук В. Ф., Крук Б. И. *Основы теории цепей*. – М.: Радио и связь, 2000. – 592 с.
2. Понтрягин Л. С. *Обыкновенные дифференциальные уравнения*. – М.: Изд-во «Наука», 1970. – 332 с.
3. Пономарёв К. К. *Составление и решение дифференциальных уравнений инженерно-технических задач*. – М.: Учпедгиз, 1962. – 184 с.
4. Попов В. П. *Основы теории цепей*. – М.: Высшая школа, 2000. – 575 с.
5. Гараев К. Г. *О некоторых проблемах математического образования в технических вузах в современных условиях* // Вестник Казанск. госу. техн. университета им. А. Н. Туполева. – 2000. – № 1. – С. 78-79.

**Е. К. Липачёв, А. С. Мохов**

*Казанский (Приволжский) федеральный университет,  
elipachev@gmail.com, alexphenomenon2010@mail.ru*

**МЕТОД ВЫЯВЛЕНИЯ ДУБЛИКАТОВ В  
РАЗНОЯЗЫЧНЫХ ЦИФРОВЫХ КОЛЛЕКЦИЯХ  
ФИЗИКО–МАТЕМАТИЧЕСКИХ ДОКУМЕНТОВ**

Одной из задач управления цифровыми коллекциями является поиск дубликатов — документов, имеющих одинаковое содержание. В случае коллекций, включающих документы с текстами на разных языках, необходимы также программные инструменты, позволяющие найти все документы, являющиеся переводами одного и того же документа. Эта задача тесно связана с задачей выявления заимствований в научных публикациях (см., напр., [1]). При этом сложнее всего найти в

тексте фрагменты, являющиеся переводом с другого языка уже опубликованных документов. Для решения этой задачи используются подходы, основанные на сопоставлении формализованных представлений документов (см., напр., [2, 3]).

В работе предложены алгоритмы поиска дубликатов в физико-математических коллекциях документов. Для каждого документа автоматически формируется “формульный код” в виде строки, учитывающей положение формул в тексте, а также структуру формул. Проведенная аналогия с задачей расшифровки ДНК позволила использовать строковые алгоритмы вычислительной молекулярной биологии [4]. Предложен алфавит “формульного кода”, система весов для формул и математических символов, таблицы стоп-символов и суффиксов [5]. Для определения близости документов используется метрика Левенштейна [6] в пространстве строк коллекции. Проведено тестирование алгоритма на коллекции документов портала Math-Net.Ru (<http://www.mathnet.ru/>). Предложена модификация алгоритмов для коллекций документов, имеющих представление формул в формате MathML (см., напр., [7]).

Работа выполнена при поддержке РФФИ (проекты №№ 15-07-08522, 15-47-02472).

## Л И Т Е Р А Т У Р А

1. Olver P. *Journals in Flux* // Notices of the AMS. – 2011. – 58 (8). – P. 1124–1126.
2. Захаров В.Н., Хорошилов Ал-др.А., Хорошилов Ал-ей.А. *Метод выявления заимствований в текстах разноязычных документов* // Аналитика и управление данными в областях с интенсивным использованием данных: труды XVIII междунар. конф. DAMDID/RCDL’2016. – М.: ФИЦ ИУ РАН, 2016. – С. 277–282.

3. Ceska Z., Toman M., Jezek K. *Multilingual plagiarism detection* // Lecture Notes in Computer Science. – 2008. – 5253. – P. 83–92.

4. Gusfield D. *Algorithms on strings, trees, and sequences: computer science and computational biology*. – Cambridge University Press, 1997. – 311 p.

5. Мохов А.С. *Метод определения подобия научных документов на основе анализа структуры формул*. – Выпускная квалификационная работа. – Казань: КФУ, 2016. – 47 с.

6. Левенштейн В.И. *Двоичные коды с исправлением выпадений, вставок и замещений символов* // Докл. АН СССР. – 1965. – 163 (4). – С. 845–848.

7. Елизаров А.М., Липачев Е.К., Малахальцев М.А. *Основы MathML. Представление математических текстов в Internet. Практическое руководство*. – М.: Физматлит, 2010. – 192 с.

**А. И. Лубышева, К. А. Поташов**

*Казанский (Приволжский) федеральный университет,  
nastia.lubysheva@gmail.com*

## **АДАПТАЦИЯ МОДЕЛИ МНОГОЗВЕННОЙ СИСТЕМЫ МИКРООРГАНИЗМОВ В ЗАГРЯЗНЕННОЙ НЕФТЕПРОДУКТАМИ ПОЧВЕ**

Наиболее экологичные методы очистки почв от нефтяных загрязнителей относятся к технологии биоремедиации, основанной на использовании углеводородоокисляющих микроорганизмов (МО). Достоверное определение оптимального сценария восстановительных мероприятий, а также прогноз масштаба загрязнения должны основываться на расчетах в рамках